**Chapter 12**

**Uncertain Knowledge and Reasoning**

**The chapter consist of Short type Questions & Answers , Descriptive Question & Answer and MCQs & answers**

Contents

[Short type Questions & Answers 1](#_Toc14943821)

[Q What is Uncertainty 2](#_Toc14943822)

[Q What are Sources of Uncertainty 2](#_Toc14943823)

[Describe the difference between marginal independence and conditional independence. 2](#_Toc14943824)

[Why are we more often interested in conditional independence? 2](#_Toc14943825)

[What are the three key components of a belief network? 2](#_Toc14943826)

[What is the independence assumption in a belief network? 2](#_Toc14943827)

[Q List Reasons for uncertainty in AI 2](#_Toc14943828)

[Q What Is Called As Decision Theory? 3](#_Toc14943829)

[Q When probability distribution is used? 3](#_Toc14943830)

[Q Define Prior probability 3](#_Toc14943831)

[Q Define Dempster-Shafer theory. 3](#_Toc14943832)

[Descriptive Question & Answer 4](#_Toc14943833)

[Q Explain Conditional probability. 4](#_Toc14943834)

[Q Explain Joint Probability with example 5](#_Toc14943835)

[Q Explain Baye’s theorem with proper case study. 6](#_Toc14943836)

[Q What is role of Normalization in uncertainty 7](#_Toc14943837)

[Q Give a short description of hidden markov models 9](#_Toc14943838)

[What do u understand from the term probabilistic reasoning 10](#_Toc14943839)

[Dental Diagnosis uncertainty, explain in detail 16](#_Toc14943840)

[MCQs & answers 17](#_Toc14943841)

# Short type Questions & Answers

## Q What is Uncertainty

Uncertainty means that many of the simplifications that are possible with deductive inference are no longer valid.

Uncertainty is unavoidable in everyday reasoning and in many real-world domains.

Examples:

• Waiting for a colleague who has not shown up for a meeting.

• Deciding whether to go to school on a very snowy winter morning.

• Judgmental domains such as medicine, business, law and so on.

## Q What are Sources of Uncertainty

Sources of Uncertainty are

Incomplete knowledge. E.g., laboratory data can be late, medical science can have an incomplete theory for some diseases.

• Imprecise knowledge. E.g., the time that an event happened can be known only approximately.

• Unreliable knowledge. E.g., a measuring instrument can be biased or defective.

## Describe the difference between marginal independence and conditional independence.

Answer: For variables X and Y , if P(X|Y ) = P(X) then X is marginally independent of Y . That is, knowing the value of Y doesn’t change the probability of X. For conditional independence, if P(X|Y, Z) = P(X|Z) then X is conditionally independent of Y given Z. In other words, if you know the value of Z, knowing the value of Y doesn’t tell you anything more about X.

## Why are we more often interested in conditional independence?

Answer: Conditional independence is more common - complete independence between variables is relatively rare in a given domain. Also, in any given domain, we typically know something, and independence queries should be conditional on that something.

## What are the three key components of a belief network?

Answer: 1. A directed acyclic graph, with each node a random variable.

2. A domain for each of the random variables.

3. A set of conditional probability distributions giving P(X|parents(X)) for each variable X.

## What is the independence assumption in a belief network?

Answer: Each random variable is conditionally independent of its non-descendants given its parents.

## Q List Reasons for uncertainty in AI

1. Laziness: it is too much work to list the complete set of antecedent or consequents needed to ensure an exception less rule and too hard to use such rules.
2. Theoretical ignorance: medical science has no complete theory for the domain.
3. Practical ignorance: even if all the rules are known, it might be uncertain about a particular patient because not all the necessary tests have been or can be run.

## Q What Is Called As Decision Theory?

Preferences As Expressed by Utilities Are Combined with Probabilities in the General Theory of Rational Decisions Called Decision Theory. Decision Theory = Probability Theory + Utility Theory

## Q When probability distribution is used?

If we want to have probabilities of all the possible values of a random variable probability distribution is used. Eg: P(weather) = (0.7,0.2,0.08,0.02). This type of notations simplifies many equations.

## Q Define Prior probability

Prior probability P(A) for the unconditional or prior probability that the proposition A is true. For example, if cavity denotes the proposition that a particular patient has a cavity, P(cavity)=0.1 .

Propositions can also include equalities involving so-called random variables. For example, if we are concerned about the random variable weather, we might have

P(weather= sunny) = 0.7

P(weather= rain)= 0.2

P(weather= cloudy)= 0.08

P(weather=snow)= 0.02

P(cavity) is P(cavity= true)

P(⌐ cavity) P(cavity= false)

P(weather) = (0.7, 0.2, 0.08, 0.02)

Q Define Dempster-Shafer theory.

The Dempster–Shafer theory (DST) is a mathematical theory of evidence. It allows one to combine evidence from different sources and arrive at a degree of belief (represented by a belief function) that takes into account all the available evidence. The theory was first developed by Arthur P. Dempster and Glenn Shafer

# Descriptive Question & Answer

## Q Explain Conditional probability.

Once the agent has obtained some evidence concerning the previously unknown random variable making up the domain, prior probabilities are no longer applicable. Instead, we use conditional or posterior probabilities. The notation used is P (a | b), where a and b are any propositions. This is read as “the probability of a, given that all we know is b”. For example,

P (cavity| toothache) = 0.8

Indicates that if a patient is observed to have a toothache and no other information is available, then the probability of patient having cavity will be 0.8. A prior probability, such as (| cavity), can be thought of as special case of the conditional probability P (cavity |), where the probability is continued on no evidence.

Conditional probabilities can be defined in terms of unconditional probabilities. The defining equation is :

P(a| b) = P (a ∧ b)/ P (b) (1)

Which hold whenever P (b) > 0. This equation can also be written as:

P (a ∧ b) = P(a| b) P(b)

Which is called he product rule. The product rule is perhaps easier to remember. It comes form the fact that for a and b to be true, we also need a to be true given b. we can also have it the other way around:

P(a ∧ b) = P(b | a) P(a)

We have defined a syntax for propositions and for prior and conditional probability statements about hose propositions. We must provide some sort of semantics for probability statements. We begin with basic axioms that serve to define the basic probability scale and its endpoints:

1. All probabilities are between 0 and 1 for proposition a,

0< P(a)< 1

1. Necessarily true proposition have probability 1 and necessarily false propositions have probability 0.

P(true)=1 P(false)=0

Next, we need an axiom that connects the logically related propositions. The simplest way to do this is to define the probability of a disjunction as follows:

1. The probability of disjunction is given as :

P(a ∨ b) = P(a) + P(b) - P(a ∧ b) (2)

This rule is easily remembered by noting that the cases where a holds, together with the cases where b holds, certainly cover all the cases where a ∨ b holds. But summing the two sets of cases counts their intersection twice, so we need to subtract P(a ∧ b).

These three axioms are often called Kolmogorov’s axioms in honor of the Russian mathematician. Andrel Kolmogorov, who showed how to build up the rest of probability theory from this simple foundation.

## Q Explain Joint Probability with example

Given random variables , that are defined on a probability space, the joint probability distribution for is a probability distribution that gives the probability that each offalls in any particular range or discrete set of values specified for that variable.

In the case of solely 2 random variables, this is called a bivariate distribution, but the concept generalizes to any number of random variables, giving a multivariate distribution.

The probability distribution will be expressed either in terms of a joint additivedistribution operate or in terms of a probability density operate (in the case of continuous

variables) or probability mass operate (in the case of distinct variables).

These successively will be accustomed notice 2 alternative sorts of distributions: the marginal distribution giving the possibilities for anyone of the variables with no relation to

any specific ranges of values for the opposite variables, and also the chancedistribution giving the possibilities for any set of the variables conditional on explicitvalues of the remaining

variables.

For example: Coin flip

Consider the flip of two fair coins; let {\displaystyle A} and {\displaystyle B} be discrete random variables associated with the outcomes of the first and second coin flips respectively. Each coin flip is a Bernoulli trial and has a Bernoulli distribution. If a coin displays "heads" then the associated random variable takes the value 1, and it takes the value 0 otherwise. The probability of each of these outcomes is 1/2, so the marginal (unconditional) density functions are

P(A)= ½ for A ε {0,1}

P(B)= ½ forB ε {0,1}

The joint probability density function of {\displaystyle A} and {\displaystyle B} defines probabilities for each pair of outcomes. All possible outcomes are

(A=0, B=0), (A=1, B=0), (A=0, B=1), (A=1, B=1)

Since each outcome is equally likely the joint probability density function becomes

P(A, B) = ¼ for A, B ε {0,1}

Since the coin flips are independent, the joint probability density function is the product of the marginals P(A, B) =P(A) P(B) for A,B {\displaystyle P(B)=1/2\quad {\text{for}}\quad B\in \{0,1\}.}ε {0,1}

## Q Explain Baye’s theorem with proper case study.

Product rule in 2 forms because of the commutativity of conjunction:

P(a ∧ b)= P(a/b) P(b)

P(a ∧ b)= P(b/a) P(a)

Equating the two right-hand side and dividing by P(a),

P(b/a) =P(a/b) P(b) / P(a)

This equation is known as Baye’s rule (Baye’s law or BAye’s theorem)

This simple equation underlies all modern AI systems for probabilistic inference. The more general case of multivalued variables can be written in the P notation as

P(y/x) = P(x/y) P(y)/P(x)

Where again this is to be taken as representing the set of equations relating corresponding elements of the table. We will also have occasion to use ore general version conditionalized on some background evidence e:  
 P(y/x, e) = P(x/y, e) P(y/e) / P(x/e)

Applying Baye’s rule: The simple case

On the surface, Bayes’ rule doesn’t seem very useful. It requires 3 terms- a conditional probability and 2 unconditional probability. Bayes’ rule is useful in practice because there are any cases where we have good probability estimates for 3 numbers and compute for 4th. In a tsk such as medical diagnosis, we often have conditional probabilities on casual relationships and want to derive a diagnosis. The doctor knows that the diseases meningitis causes cause the patient to have stiff neck, say, 50% of the time. The doctor also knows some unconditional facts : the prior probability of patient having meningitis is 1/50,000 , and the prior probability of patients having stiff neck is 1/20. Letting s be the proposition that patient has stiff neck and m be the proposition that patient has meningitis,

We have, P(s/m) =0.5

P(m) =1/50,000

P(s) = 1/20

P(m/s) = P(s/m) P(m) / P(s)

= 0.5\* 1/50000 / (1/20)

= 0.0002

That is, we expect only 1 in 5000 patients with stiff neck to have meningitis. Notice that even though stiff neck is strongly indicated by meningitis (with probability 0.5), the probability of meningitis in the patient remain small. This is because the prior on stiff neck is much higher than that for meningitis.

## Q What is role of Normalization in uncertainty

Consider the equation for calculating the probability of meningitis given a stiff neck:

P(m/s) =P(s/m) P(m) / P(s)

Suppose we are also concerned with the possibility that the patient is suffering from whiplash w given a stiff neck:

P(w/s) = P(s/w) P(w) / P(s)

Comparing these 2 equations, we see that in order to compute the relative likelihood of meningitis and whiplash, given a stiff neck, we need not assess the prior probability P(s) of a stiff neck. To put numbers on the equations, suppose that

P(s/w) =0.8 and P(w) =1/ 1000 Then

P(m/s)= P(s/w) P(m) ~ 0.5 \*1/50000 ~ 1   
P(w/s) =P(s/w) P(w) ~ 0.8 \* 1/1000 ~ 80

That is, whiplash is 80 times more likely than meningitis, given a stiff neck. In some cases, relative likelihood is sufficient for decision making, but when , as in this case, the two possibilities yield radically different utilities for various treatment actions, one need exact values to make rational decisions. It is still possible to avoid direct assessment of the prior probability of “symptoms”, by considering an exhaustive set of cases .

For example, we can write equation for m and ⌐ m

P(m/s) =P(s/m) P(m) / P(s)

P(⌐ m/s) =P(s/⌐m) P(⌐m) / P(s)

Adding these 2 equations, and suing the fact that

P(m/s) + P(⌐ m/s) =1, we obtain

P(s) =P(s/m) P(m) + P(s/⌐m) P(⌐m)

Substituting into the equation for P(m/s), we have

P(m/s) = P(s/m) P(m) / P(s/m) P(m) + P(s/⌐m) P(⌐m)

This process is called normalization, because it treat 1/ P(s) as a normalizing constant that allows the conditional terms to sum to 1. Thus in return for assessing the probability P(s/⌐m) we can avoid assessing P(s) and still obtain exact probabilities from Bayes’ rule.

The general form of Bayes’ rule with normalization is

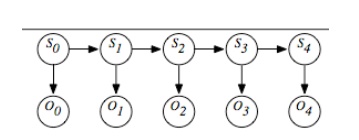
P(y/x) =α P(x/y) P(y)

Where α is the normalization constant needed to make the entries in P(y/x) sum to 1. The usual way to use normalization is to calculate the unnormalized values, and then scale them all so that they add to 1.

## Q Give a short description of hidden markov models

A hidden Markov model (HMM) is an augmentation of the Markov chain to include observations. Just like the state transition of the Markov chain, an HMM also includes observations of the state. These observations can be partial in that different states can map to the same observation and noisy in that the same state can stochastically map to different observations at different times.

The assumptions behind an HMM are that the state at time t+1 only depends on the state at time t, as in the Markov chain. The observation at time t only depends on the state at time t. The observations are modeled using the variable for each time t whose domain is the set of possible observations. The belief network representation of an HMM is depicted in Figure. Although the belief network is shown for four stages, it can proceed indefinitely.



A stationary HMM includes the following probability distributions:

P(S0) specifies initial conditions.

P(St+1|St) specifies the dynamics.

P(Ot|St) specifies the sensor model.

There are a number of tasks that are common for HMMs.

The problem of filtering or belief-state monitoring is to determine the current state based on the current and previous observations, namely to determine P(Si|O0,...,Oi).

Note that all state and observation variables after Si are irrelevant because they are not observed and can be ignored when this conditional distribution is computed.

The problem of smoothing is to determine a state based on past and future observations. Suppose an agent has observed up to time k and wants to determine the state at time i for i<k; the smoothing problem is to determine

P(Si|O0,...,Ok).

All of the variables Si and Vi for i>k can be ignored.

## What do u understand from the term probabilistic reasoning

                    Representing Knowledge in an Uncertain Domain

                    Belief network used to encode the meaningful dependence between variables. o Nodes represent random variables

o  Arcs represent direct influence

o Nodes have conditional probability table that gives that var's probability given the different states of its parents

o  Is a Directed Acyclic Graph (or DAG)

**The Semantics of Belief Networks**

                    To construct net, think of as representing the joint probability distribution.



                    To infer from net, think of as representing conditional independence statements.



                    Calculate a member of the joint probability by multiplying individual conditional probabilities:



o  P(X1=x1, . . . Xn=xn) =

o  = P(X1=x1|parents(X1)) \* . . . \* P(Xn=xn|parents(Xn))

                    Note: Only have to be given the immediate parents of Xi, not all other nodes:

o  P(Xi|X(i-1),...X1) = P(Xi|parents(Xi))

                    To incrementally construct a network:

1.                             Decide on the variables

2.                             Decide on an ordering of them

3.                             Do until no variables are left:

a.                                                     Pick a variable and make a node for it

b.                                                     Set its parents to the minimal set of pre-existing nodes

c.                                                      Define its conditional probability

                    Often, the resulting conditional probability tables are much smaller than the exponential size of the full joint



                    If don't order nodes by "root causes" first, get larger conditional probability tables

                    Different tables may encode the same probabilities.

                    Some canonical distributions that appear in conditional probability tables:

o deterministic logical relationship (e.g. AND, OR) o deterministic numeric relationship (e.g. MIN)

o parameteric relationship (e.g. weighted sum in neural net) o noisy logical relationship (e.g. noisy-OR, noisy-MAX)

**Direction-dependent separation or D-separation:**

                    If all undirected paths between 2 nodes are d-separated given evidence node(s) E, then the 2 nodes are independent given E.



                    Evidence node(s) E d-separate X and Y if for every path between them E contains a node Z that:



o has an arrow in on the path leading from X and an arrow out on the path leading to Y (or vice versa)

o  has arrows out leading to both X and Y

o  does NOT have arrows in from both X and Y (nor Z's children too)

**Inference in Belief Networks**

                    Want to compute posterior probabilities of query variables given evidence variables.



                    Types of inference for belief networks:

o Diagnostic inference: symptoms to causes o Causal inference: causes to symptoms

o                 Intercausal inference:

o                 Mixed inference: mixes those above

**Inference in Multiply Connected Belief Networks**

                    Multiply connected graphs have 2 nodes connected by more than one path

                    Techniques for handling:

o                 Clustering: Group some of the intermediate nodes into one meganode. Pro: Perhaps best way to get exact evaluation.

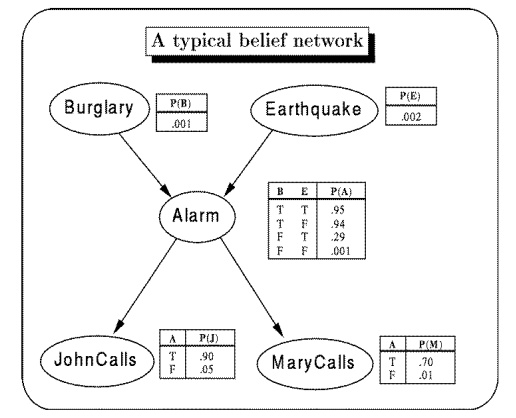
Con: Conditional probability tables may exponentially increase in size.

o                 Cutset conditioning: Obtain simplier polytrees by instantiating variables as constants.

Con: May obtain exponential number of simplier polytrees.

Pro: It may be safe to ignore trees with lo probability (bounded cutset conditioning).

Stochastic simulation: run thru the net with randomly choosen values for each node (weighed by prior probabilities).



## Dental Diagnosis uncertainty, explain in detail

To act rationally under uncertainty we must be able to evaluate how likely certain things are. With FOL a fact F is only useful if it is known to be true or false. But we need to be able to evaluate how likely it is that F is true. By weighing likelihoods of events (probabilities) we can develop mechanisms for acting rationally under uncertainty.

**Dental Diagnosis example.**

In FOL we might formulate

P. symptom(P,toothache)→  disease(p,cavity)              disease(p,gumDisease)

disease(p,foodStuck)

When do we stop?

Cannot list all possible causes.

We also want to rank the possibilities. We don’t want to start drilling for a cavity before checking for more likely causes first.

**Axioms Of Probability**

Given a set U (universe), a probability function is a function defined over the subsets of U that maps each subset to the real numbers and that satisfies the Axioms of Probability

1.Pr(U)  = 1

2.Pr(A)  [0,1]

3.Pr(A   ∈B) = Pr(A) + Pr(B) –Pr(A ∩B)

∪

Note if A ∩B = {} then Pr(A ∪B) = Pr(A) + Pr(B)

# MCQs & answers

**1.** Using logic to represent and reason we can represent knowledge about the world with facts and rules.

(a) True

(b) False

**2.** Uncertainty arises in the wumpus world because the agent’s sensors give only\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

(a) Full and global information

(b) Partial and global information

(c) Partial and local information

(d) Full and local information

**3.** How is Fuzzy Logic different from conventional control methods?

(a) IF and THEN approach

(b) FOR approach

(c) WHILE approach

(d) DO approach

**4.** What is used for probability theory sentences?

(a) Conditional logic

(b) Logic

(c) Extension of propositional logic

(d) None of the above

**5.** Where does the dependence of experience is reflected in prior probability sentences?

(a) Syntactic distinction

(b) Semantic distinction

(c) Both Syntactic and Semantic distinction

(d) None of the above

**6.** What is the basic element for a language?

(a) Literal

(b) Variable

(c) Random variable

(d) All of the above

**7.** What is meant by probability density function?

(a) Probability distributions

(b) Continuous variable

(c) Discrete variable

(d) Probability distributions for continuous Variables

**Answers**

**1. (a) 2. (c) 3. (a) 4. (c) 5. (a) 6. (c) 7. (d)**